FOOD CONSUMPTION AND EXPENDITURE DATA

United States Department of Agriculture
Office of International Cooperation and Development

SAMPLING METHODOLOGIES FOR MORE COST-EFFECTIVE COLLECTION OF FOOD CONSUMPTION AND EXPENDITURE DATA

By:

Tom Zalla
Agricultural Economist

April, 1988

Report prepared for the Nutrition Economics Group, OICD, United States Department of Agriculture, Washington D.C.

Oct 28
[illegible]

Oct 31
comments
on Jones

SAMPLING METHODOLOGIES FOR MORE COST-EFFECTIVE COLLECTION OF FOOD CONSUMPTION AND EXPENDITURE DATA

951035

CONTENTS

I. INTRODUCTION

The Nutrition Economics Group (NEG) of the United States Department of Agriculture assists USAID missions throughout the world in preparing, executing and analyzing surveys of food consumption and expenditures. Consumption and expenditure surveys provide data essential for detailed analysis of a wide range of policies relating to food, nutrition, prices and incomes. The general thrust of NEG's work has been to finance innovative, cost-effective methodologies that can significantly reduce the cost and time required to gather the kind of data needed for policy analysis. The first stage of this effort has been to draw upon the efforts of a number of consultants experienced in various aspects of food consumption and expenditure analysis, each using a more or less unique methodology in a particular country.

In an attempt to capitalize on the experience of these innovative efforts by individuals with different perspectives, the NEG sponsored a workshop on Rapid Appraisal Techniques For Food Policy Analysis in April, 1987. The purpose of the workshop was to explore how various sampling and data collection methods might reduce the amount of data collected and speed their release, while retaining their usefulness for rigorous quantitative analysis. Following the workshop NEG commissioned more in-depth studies of some of the key proposals made. This report is one of those studies.

This report reviews sampling methodologies used in several consumption surveys with rapid appraisal objectives. It also summarizes some data collection and sampling methodology issues raised at the workshop. It proposes an approach to sampling for consumption surveys that is widely understood and which admits of fairly rapid application, yet which can accommodate the most rigorous of conventional techniques. It provides formulas for estimating variances that, though slightly biased in some instances, are simple to use where computers or computer programming skills are not available. The approach throughout is on realistic as-

sessment of all types of potential error in obtaining consumption and expenditure data. The objective is to make the shortcuts suggested more palatable to those who are more disciplinary in their approach to sampling.

II. THE RAPID APPRAISAL PERSPECTIVE AS APPLIED TO CONSUMPTION AND EXPENDITURE SURVEYS

One conclusion that emerged from the Rapid Appraisal Workshop is that there are significant limits on the number of shortcuts one can take and still obtain data of sufficient quality for quantitative food policy analysis. The keynote paper for the workshop (Zalla, 1987) contained a review of several non-formal and more qualitative techniques for gathering food consumption and expenditure data. Such techniques can frequently yield very good descriptive data. They also can generate an understanding of consumption/expenditure systems that is not easily obtained from questionnaire surveys.

Such techniques, however, are of limited use where quantitative analysis of food consumption interactions is required. Seasonal production and consumption patterns, the presence of lumpy income, consumption and expenditure events and the need for reasonably precise measures of consumption and expenditures for many types of analyses limit the amount of detailed data collection one can dispense with, except in very narrowly defined situations. For this reason we use the term cost-effective rather than rapid appraisal throughout this report. Nonetheless, there is considerable scope for improving the turnaround time of consumption and expenditure data, while, at the same time, reducing the cost of acquiring them. The sampling and data collection methodologies described below are directed to these ends.

III. SAMPLING METHODOLOGIES USED IN OTHER CONSUMPTION/EXPENDITURE SURVEYS

A. Madagascar Rice Demand Study

AIRD (Ahlers, 1983) assisted the Ministry of Agriculture of Madagascar in conducting a survey of rice consumption in the country's seven regional capitals in 1984. Using Ministry of Agriculture statistical and survey staff, survey organizers utilized a stratified, two-stage sample of 280 households in Antanarivo and 80 households in each of the six regional capitals. Each city in the sample was treated as a separate stratum. Antanarivo, the capital, was further stratified into income groups by means of housing type.

Local level administrative units, the fokotany, served as the primary sampling units (psu) at the first stage. The population of each fokotany was readily available from administrative sources. A cumulative listing of all households in each fokotany within a stratum was used to draw a systematic sample of fokotany with probabilities proportional to estimated size. Ten fokotany were selected in each of the smaller centers and 28 were selected in Antanarivo.

A list of the names of the heads of household living in each of the selected fokotany served as the second-stage sampling frame. As at the first stage, a systematic sampling procedure was used to ensure a representative distribution of the sample units. Sample size was constant for each fokotany within a given stratum.

A systematic sampling procedure using a cumulative population listing amounts to sampling with probability proportional to a measure of size. Each fokotany has a probability of being selected that is equal to its population relative to the population of all fokotany within the stratum. At the second stage each household has a probability of selection equal to the size of sample drawn from the fokotany divided by its household population. When used in conjunction with a sample of constant size within each selected primary sampling unit, this procedure gener-

ates a self-weighting sample. The advantage of a self-weighting sample is that it can be treated as a simple random sample for estimating population means. This can be important where computers are not available, where circumstances do not permit quick calculation of variances estimated by complex formula.

The Madagascar survey consisted of a single visit to each selected household over one six-week period of the year. Since the sample population was an urban population, researchers felt they could obtain an adequate representation of average annual household consumption with such a methodology.

B. Liberia Urban Food Consumption Survey

In conjunction with the Nutrition Economics Group of USDA Purdue University conducted a study of urban food consumption patterns in Liberia in 1986 (Heimstra, 1987). The study was limited to Monrovia, several other major urban centers and a representative number of smaller urban centers in rural areas with heavy concentrations of rice production. In total, seven urban areas were studied.

The sampling methodology utilized a type of area sampling. Each town was divided into three (six in Monrovia) sectors of roughly equal household population. These sectors were essentially geographically defined substrata within the particular urban center. Each sector/substratum was assigned to one enumerator. Enumerators prepared a sketch showing all streets and major paths in their assigned sector. They then proceeded to subdivide the sector into sub-areas of approximately 100 structures. These sub-areas served as the primary sample units.

To draw the sample enumerators counted the number of structures in each sub-area and prepared a cumulative listing of structures. A number between one and the total number of structures in the sector/substratum was selected from a random number table and the sub-area containing that structure became the selected first-stage sample unit. There was, thus, only one first-stage

unit selected per substratum. Each city was, in addition, a separate stratum.

Once the sub-areas/primary sample units were selected, enumerators listed all of the structures within the sub-area. This list of structures became the sampling frame at the second stage and the structures themselves were the second-stage sample units. A random sample of 37 structures (46 in Monrovia) was selected for interviewing. This represented roughly a 1/3 (plus 10% reserve) sample of the structures within the selected sub-areas.

For the purposes of this study, a structure was defined as one being lived in by private individuals. Apartment units that appeared to be independent housing units were defined as separate structures even though they were physically connected. Within each chosen structure enumerators selected a single household to be interviewed. Enumerators selected households by means of a predetermined alphabetical selection based on the first names of the household heads. Only households defined as a "housekeeping household", i.e. one that prepared at home at least half of its meals for own consumption during the survey period, were included in the sample.

Since the second-stage sample unit, the structure, had an unknown number of households, the <u>ex-ante</u> extrapolation weight was multiplied by the number of households found in the structure at the time of the interview. In effect, the interviewed household was assumed to be representative of all households within the structure. This had the effect of increasing the representation of smaller households in proportion to their incidence in the population of the sub-area. This was necessary in order to obtain representative household-level estimates at the level of the sub-area. For the purposes of making mean estimates for households within an urban area, households from the three (six in Monrovia) sub-areas in each urban center were combined as though they were drawn from a weighted, simple random sample . Strata weights assigned for making overall population estimates then re-

flected only the size of the individual urban centers as a proportion of the total population of all the centers combined.

The small number of first-stage sample units presented a problem in the Liberia study. Urban housing patterns tend to reflect concentrations of similar structures and household economic status. Since each sector was already being divided into sub-areas of 100 structures, it seems that it would have been relatively easy to sample one third to one fourth as many structures in three or four separate sub-areas rather than concentrating the sample in only one.

C. Haiti Household Expenditure And Consumption Survey

An expenditure survey being conducted in Haiti during 1986-87 utilized several methodological variants that offer promise with respect to rapid appraisal. The survey utilized a stratified two-stage design to select the household units to be interviewed. Unlike the other two surveys, the Haiti survey included both urban and rural areas.

To accommodate the variation in consumption patterns between urban and rural areas, and between geographically distinct rural areas in the country, the survey design divided the country into nine relatively homogeneous strata. This first level of stratification corresponded to the major geographical domains: urban and rural areas for each of the four planning regions of the country, plus a separate stratum for metropolitan Port-au-Prince.

Census enumeration areas (sections d'enumeration), were used as the primary sampling units. All enumeration areas (EA) were delineated on census sketch maps. Their boundaries normally follow defined topographical characteristics so that they can be identified on the ground. Each EA was classified as urban, rural or peri-urban. Since the peri-urban areas are generally more rural in nature, these were included in the rural strata. In addition, the urban geographic strata were subdivided into three economic substrata reflecting high, medium and low levels of in-

come. The rural strata were subdivided into plains and mountain substrata. Each EA was allocated to that economic substratum which appeared to contain most of its population.

Once EAs were classified and placed into the appropriate substratum they were listed in serpentine fashion in order to provide implicit stratification. Their household populations from the preliminary census tabulations were used as a measure of size. These measures of size were cumulated within a substratum. The total number of households in the substratum, divided by the desired number of EAs to be sampled, defined the systematic sampling interval used to select the specific EAs. Selection of households proceeded in similar fashion with 10 households sampled per EA. Since a constant size sample of 10 households per EA was selected, the sample was approximately self-weighting. In addition, since the sample within each stratum was allocated proportionately among the substrata, the sample was also approximately self-weighting within a stratum.

Because policy makers wanted more or less equally reliable estimates from each major geographic domain, the overall sample was not allocated in proportion to the populations of the strata. This prevented the sample from being completely self-weighting. The Haiti survey allocated approximately 2/3 of the sample to the urban strata based on anticipated variability. This was expected to generate more efficient estimates than would have been likely with a completely self-weighting sample, assuming the variance assumptions were correct.

Actual data collection for the Haiti survey was to be spread over a 52 week period. The selected EAs were divided into 13 subsamples, each randomly assigned to one four-week period of the year. Data was collected from selected households only once. Rotation of the sample over the year provided the necessary seasonal dimension to average consumption patterns.

When first-stage sample units are selected using non-current population figures, as is often the case when using census enumeration areas as psus, the number of households found at the time of listing will often be different from that attributed to the psu at the time of sampling. The individual household weights need to be adjusted to reflect this difference. They also must be adjusted to reflect the presence of multiple households not separately listed, consolidated households that were listed separately, as well as unoccupied and other ineligible sample housing units. The Haiti survey adjusted weights for all of these factors. Non-response adjustments were also necessary in certain EAs where fewer than 10 household interviews were achieved after all substitutes were exhausted. The final household weights then became a permanent part of the data record for each household. This did not present a computational problem since data are being analyzed by computer.

D. Kilimanjaro Food Consumption Survey

The author carried out a study of food consumption in rural Kilimanjaro District of Tanzania in 1973-4. The study utilized a 24 hour recall of all meals prepared and all food consumed by household members and guests. The sampling methodology was similar to that utilized in the Haiti survey, except that stratification was implicit rather than explicit.

Tanzania had detailed published statistics by census enumeration area and had district maps and individual sketches showing the location of each EA. Because of topographical influences on income and consumption patters, the sample area was divided into 12 contiguous vertical slices/sectors of the mountain, each constructed so as to contain essentially the same number of households. The EAs were listed in serpentine fashion beginning at the top of the first EA, proceeding to the bottom, and onto the bottom of the next one, proceeding to the top. A systematic sample of 36 EAs was drawn using 1/36 of the cumulative household population listing as the systematic sampling interval. Households were listed within each of the selected EAs and eight were

subsampled in similar fashion.

The process of selecting households with probability proportional to a measure of size at the first stage, as is accomplished with the systematic sampling procedure, coupled with a constant size sample per EA at the second stage created an essentially self-weighting sample. This was necessary since preliminary tabulations were manual. No adjustment was made for the difference between the measure of size used to draw the sample and the number of households actually listed by enumerators. In some cases these differences were substantial. However, it frequently was not clear that such differences were due to changes in the household populations of the sampled EAs as opposed to failing to identify the actual boundaries utilized in the census.

Conceptually the selected first-stage unit was only an interval on a continuum and it made little practical difference that we may have not identified the proper number of households at the time of listing the households in the field. Only if out-migration were substantial and varied at different localities on the mountain would the differences in sampled and listed populations represent a problem of practical significance. Moreover, even if such were the case, the relatively small sample size (8*36=288) meant that sampling errors would have swamped the effect of differential out-migration.

The self-weighting nature of the sample allowed use of simple random sampling estimators without introducing bias. For calculating variances the study used both a simple random sample estimator and a modified two-stage estimator given by Yates (1960, p.197) that, though slightly biased, is computationally simple. The two-stage estimator was used only on a few key variables to determine the average difference between the two methods. Variances calculated via the simple random sampling formula were then roughly corrected for the average difference between the two. This approach is adequate in rural areas where sampling errors are usually insignificant as compared to non-sampling errors.

IV. LESSONS FOR DESIGNING COST-EFFECTIVE SAMPLING METHODOLOGIES

In examining the merits of sampling methodologies from the perspective of cost-effectiveness, it is important to consider the potential for both sampling and non-sampling errors in consumption surveys. This is especially true in rural areas where household production is an important part of consumption and expenditures. It is equally true where the local agencies responsible for implementation lack manpower, material resources, work discipline or adequate supervision.

Sampling errors arise both from faulty sample design and from the fact that not all members of the population are interviewed. Non-sampling errors arise from such factors as poorly trained enumerators, poorly designed and tested survey instruments and survey operations, untimely field support to correct problems in execution, lack of supervision to ensure enumerators are performing to standard, imprecise measures or recall periods, poor translation, respondent deception or lack of knowledge and a host of other factors. In the typical LDC context errors arising from these kinds of factors are usually quantitatively much more significant than sampling errors. They also increase with sample size. From the point of view of accuracy of estimates, survey designers need to balance resources and time spent constructing, utilizing and making inferences from an unbiased sampling frame with resources and time devoted to field data collection. The former minimizes required sample size for a given level of precision while the latter maximizes data quality and improves accuracy.[1]

---

[1] Precision refers mainly to sampling error, i.e. the extent to which a sample approximates the underlying data from which it is drawn. Accuracy refers to precision plus bias. Many very useful estimators are slightly biased. They are used where the bias is small because they are simple to employ or are practically measurable. Bias includes mostly the effect of non-sampling error. Non-sampling error is largely an unknown element, though we can estimate certain aspects of it.

There is also a relationship between the length of the questionnaire and data quality that interacts with sample size. For a given budget a larger questionnaire will permit a smaller sample size. The appropriate trade-off will depend on survey objectives and researcher perception of the gains or losses in quality as these relate to both the size and the layout of the survey instrument and the frequency of visits to survey households. These are as much problem identification, data collection and questionnaire design issues as they are sampling design issues. Sampling methodology will, in most cases, be dictated by how researchers perceive these other issues.

In general terms, a self-weighting sample provides greater flexibility with respect to rapid tabulation of results under the widest set of circumstances. Self-weighting samples reduce to simple random samples for calculating totals and means. This is very helpful where manual tabulation is necessary or computer programing skills are limited. As was done with the Kilimanjaro survey, variances can be approximated by first calculating a dozen or so variances using both the simple random sample formula and a slightly biased but easy to calculate two-stage formula. The remaining variances can be calculated using the simple random sample formula, correcting for the proportional difference between the two methods as suggested by the results of those actually calculated using both methods. Of course, where facilities and resources permit, the same sampling methodology can accommodate variable weighting for a more accurate, less biased result.

The Haiti and Kilimanjaro surveys provide examples of various aspects of a suitable rapid appraisal sampling methodology. Both surveys employed listing and stratification techniques of one sort or another to improve sample efficiency. Both surveys sampled with probability proportional to size; and both collected data from 25-35 first-stage units and utilized a constant sized sample in the selected cluster areas. These procedures generated samples that were largely self-weighting within a stratum and included a sufficient number of psus to provide reasonably precise estimates.

These are the essential ingredients for a flexible, efficient and easily prepared, executed and tabulated sample survey.

The Haiti survey provided a novel methodological variant that has wide applicability in politically unstable environments. Dividing the sampled EAs into 13 randomly selected subsamples meant that each subsample provided an independent estimate of national consumption, albeit one subject to considerable seasonal bias. This proved valuable when political unrest forced cessation of data collection after collecting only eleven months of data. In spite of collecting less data than planned, the survey was complete for the nation. Had survey planners chosen to survey in blocks requiring the completion of all 13 four-week periods to obtain a complete sample, the usefulness of data which were actually able to be collected would have compromised.

The decision to interview each household only once and to use separate households for measuring consumption and expenditures at different times of the year is another interesting methodological variant of the Haiti survey. The researchers argue convincingly that this approach is more rapid and cost-effective where analytical needs can be satisfied with data for typical groups of households rather than for individual households. Accurately measuring nutrient intake, in particular, requires many more visits per household than those commonly recommended for repeated household visits in surveys in developing countries (Freudenheim et al., 1987). By settling for data for population groups the total number of interviews required, and thus the cost of the survey, is significantly reduced.

In the typical rapid appraisal context researchers will have to rely on administrative lists of population or dated census information for a sampling frame. All of the methodologies reviewed above utilized one of the two. Administrative lists of households, used in Madagascar and Liberia, are generally more available and more clearly defined. The smallest subdivisions for which population data can be found in the capital city can serve

as first-stage sample units. Maps of census enumeration areas, especially in rural areas are frequently nonexistent or lack sufficient detail to permit precise re-identification on the ground. But the census enumeration areas have the advantage of being small, of similar size and well distributed across ecological and agricultural variation - two factors having a profound impact on consumption patterns. Ideally, a rapid appraisal sampling methodology should be robust enough to adapt itself readily to either type of sampling frame.

Administrative and census population data usually give a reasonably good measure of the actual size of the administrative or census enumeration areas to be included in the sampling frame. Since some households may have moved or doubled up since the last data were collected, it is always important to verify the sampling methodology's ex ante estimate of the size of the sample units with what enumerators find at the time of actually listing households. Weights can then be assigned to the individual households surveyed to correct for large discrepancies.

In a relatively stable context, such as existed in Kilimanjaro, or where the measure of size (previous census or administrative statistics) is recent, one could sometimes ignore such adjustments without affecting the results appreciably. A quick comparison of the listed population with the census figures will reveal whether such an adjustment is necessary. Much will depend on how the data are to be tabulated. Unless done by computer, tabulating data from weighted households can be very tedious indeed. Ignoring minor differences between the measure of size used in sampling and the actual size found at listing allows each observation to have an equal weight within the sub-population in question. As a consequence, except for computing variances, the data can be treated as a simple random sample.

But where the differences are significant, as they frequently are in urban areas, and where those differences appear to arise from population dynamics rather than identification or listing

problems, some adjustment needs to be made. Otherwise, the survey measures that which existed at the time the population data were collected rather than what exists at the time of the consumption survey. Significant migration of young heads of household seeking urban employment or settling new lands are common situations in Africa that give rise to a need for adjusting sampling weights in light of what enumerators find at the time of listing.

Turning to the question of overall sample size, what is appropriate depends upon what policy makers intend to do with the data, the amount of variability in the population with respect to the kind of data being collected, sampling methodology and resource and operational constraints. Megill and Dauphin (1986) of the US Bureau of the Census indicate that experience with consumption surveys in LDCs suggests that in countries with population distributions similar to Haiti, 3000 households should be sufficient to provide reliable urban and rural estimates at the national level as well as estimates of predominant characteristics at the regional level. The study in Liberia got good results for urban areas with 942 households. But one would not expect to obtain the same results in Zaire or in some Asian countries where consumption patterns vary significantly between regions. Where a previous consumption survey exists variance in that sample can be used to more precisely determine an appropriate sample size for the current survey.

In most surveys it is desirable for operational reasons to use multistage sampling. This greatly reduces the cost of identifying a sample and interviewing the selected households. Survey work is more concentrated and quality can be more easily controlled. As a result, non-sampling errors are usually considerably smaller.

On the other hand, multistage sampling introduces an element of similarity between sample units that is usually greater than in the population at large. This arises from people's tendency to live near people more like themselves. The higher the correlation

between sample units within a cluster, the more precision is lost by concentrating a given sample in fewer clusters.

Megill and Dauphin (1985) indicate that studies of intraclass correlation done for consumption surveys in general suggest that the range of optimum values for the number of sample households per cluster normally includes 10. This was the number used in the Haiti survey. The Kilimanjaro survey used eight per cluster (Zalla, 1982). There, eight households in 36 first-stage sample units in a relatively homogeneous area produced coefficients of variation for estimated means for universally distributed nutrition variables that averaged 2.5% when the data were treated as a simple random sample and 2.8% when treated correctly as a two-stage sample. With commodity by commodity consumption data the standard errors would be considerably larger. And the relative differences between the simple random and two-stage estimates would be larger in less homogeneous areas. But these data support the general conclusion that sampling 8-10 households per cluster provides a good balance between providing sufficient first-stage sample units to generate reasonably precise estimates on the one hand, and concentrating data collection efforts in order to control data quality and minimize costs, on the other.

Taking the mean values of 36 first-stage units to calculate a variance for a two-stage estimate is not the same as using only 36 observations. The first-stage means already average out a great deal of variation in the data. Only where each cluster of sample units was very homogeneous within a cluster and very heterogeneous between clusters would the intraclass correlation be so high as to greatly reduce the precision obtained from a multistage sample versus a simple random sample of the same number of elements as psus in the multistage sample. Even in these cases the population can often be divided into strata so as to maximize variation between the strata and minimize variation between first-stage sample units. Such stratification sometimes more than offsets the effect of clustering within a stratum. Stratification is frequently quite feasible in urban areas where housing patterns are a good

proxy for consumption patterns.

What this all means is that there is not usually a great conflict between the need to identify, interview and tabulate sample elements rapidly, and the need to have results that reasonably accurately reflect food consumption patterns within the sample area. The major additional cost of a more rigorous and statistically sound sampling methodology is the time and money required to list all sample units within a greater number of sample clusters and to make whatever social and political introductions are necessary to enlist the cooperation of sample households within each of the selected clusters. Transportation time and costs will also be greater but they need not slow the collection of data appreciably over a concentrated sample population. Preparation of the sample frame and actual collection, tabulation and analysis of the data are not affected in a significant way. This is where the greatest portion of time and money is normally spent.

V. THE NEG RAPID APPRAISAL WORKSHOP

The NEG April, 1987 workshop on Rapid Appraisal Techniques For Food Policy Analysis included an entire session on sampling issues. The discussion centered around the importance of minimizing non-sampling errors in data collection; the one-visit, staggered approach to data collection used in the Haiti survey; and the tendency for secondary users of survey data to analyze the data as though it were collected using a simple random sample methodology. Most of the issues relating to benefits of stratification and multistage sampling mentioned above were also discussed.

As with many of the sessions, several of these issues were raised without moving to a resolution or a position being taken. This was to be the purpose of the afternoon working-group discussions. These, in turn, were too short for what had to be done. As a result, the afternoon working sessions did not really

accomplish much more than greater depth of discussion of issues raised in the plenary session. For this reason one cannot really say that the sampling methodology for rapid collection of consumption and expenditure data recommended in this paper reflects a consensus of opinion among the data collection methodology discussion group. At the same time, however, little in the methodology conflicts with what was said, either.

The one point about which there was some disagreement was the desirability of sampling so as to obtain a self-weighting sample. One of the participants felt this was too restrictive and unnecessary. It usually requires cutting corners that, in some circumstances may not be wise. This certainly is a valid concern where the data on which the sampling frame is based are out of date. For this reason the recommended sampling methodology facilitates a self-weighting sample, but does not require it.

One of the more interesting ideas discussed at the workshop was the methodology being used in Haiti to gather data. As was mentioned previously, the selected first-stage sample units were divided into 13 systematically selected subsamples. One of the subsamples was interviewed every four weeks. Each selected household was thus interviewed only once.

Though workshop participants did not reach any obvious conclusion regarding the merits or disadvantages of the single interview versus the multiple interview approach, many of the points raised in the keynote paper by the author (Zalla, 1987) were mentioned. These included the ease of revisiting a previously interviewed household versus the probability of frequent absence or non-response of a sampled household on subsequent interviews because of the unusual mobility of Haitians. There was some skepticism among participants that multiple interviews of the same household increase cooperation and data quality as much as is frequently assumed. Nor, in Haiti, does it seem necessary to spend a great deal of time informing and gaining approval of respondents for conducting the interview. The major cost there is the cost of

conducting an actual household interview. Since the within household correlation between consumption or expenditures at different periods of time is probably quite high, many fewer actual household interviews are required to obtain a given level of precision with the single interview approach. All of this assumes, of coarse, that the analytical objectives do not require data on individual households. This is a common situation for expenditure surveys but is less true for consumption surveys.

In a context where data are collected on an individual household more than once, the sampling methodology effectively includes another stage. The last stage is household consumption or expenditures over time. More than one observation gives a better measure, caeteris paribus, than does a single observation. In effect, sample size at this last stage is equal to the number of households surveyed times the number of interviews per household. A single interview per household for a sample of given size substantially reduces effective sample size. Though it would depend on just how stable consumption and expenditure patterns within a given household are over time, a single interview sample would need to contain a larger number of households than a multiple interview sample to obtain an equal level of precision. It would, however, generally require far fewer total interviews.

The increase in sample size required to obtain an equal level of precision depends on the magnitude of the intraclass (i.e. intrahousehold) correlation between consumption or expenditures in one period and that in another. This intraclass correlation will probably be higher in urban areas than in rural areas. It will probably be higher in very wealthy and very poor households as well. It will also probably be higher for expenditures than for consumption. The higher this correlation, the greater will be the gains in precision from a single interview to a larger sample. These are measurable phenomena; and this is an important issue for NEG. It bears directly on rapid appraisal data collection methodology where individual household level data is not required.

Before one can make a definitive pronouncement concerning the potential savings in cost and time of the single-interview, time-phased data collection technique used in Haiti, NEG will need to finance a combination consumption/expenditure survey that includes an examination of both the design effect and the effect on data quality arising from interview frequency. This is an issue quite distinct from the sampling methodology suggested in this report. It concerns primarily optimal sample size for the two approaches to data collection. The sampling methodology itself can accommodate virtually any approach to data collection.

VI. RECOMMENDED SAMPLING METHODOLOGY FOR NEG SURVEYS

Because of the wide variation in circumstances under which this sampling methodology will be applied, it is desirable that it be statistically rigorous while admitting shortcuts that have a more or less determinable effect on the precision and accuracy of the estimates. A good way to accomplish this is with a multistage stratified sample that is largely self-weighting. Multistage sampling reduces the cost of data collection but creates a design effect that reduces the precision of estimates vis a vis simple random sampling. Appropriate stratification, on the other hand, improves efficiency over simple random sampling and helps offset the design effect of multistage sampling.

A. Sampling Frame And Sampling Procedures

A list of census enumeration areas or of the lowest level administrative units and their respective sample unit populations makes a readily accessible and easy to use first-stage sampling frame for consumption surveys in most countries. These first-stage units can usually be grouped into strata and substrata of known population and economic or geographical characteristics. They are usually small enough that one enumerator can cover the area on foot.

The most frequent criteria used to stratify sample units in consumption surveys are geographical (urban versus rural), ecological (as these relate to major differences in agricultural production patterns) and economic (large differences in income as indicated by housing patterns, location, ethnic group or race). Strata and substrata are defined according to an hierarchy of importance _vis a vis_ analytical objectives and the most logical sampling procedure. Ideally, the substrata should be defined such that proportional allocation of the sample to the various substrata within a strata will provide a reasonably efficient sample design. This can limit necessary weighting to the stratum means, reducing substantially the aggregation burden for manually tabulated surveys. Moreover, the strata should be defined so as to maximize the variation between the strata and minimize the variation within a stratum. This improves the precision of the estimates over those from a simple random sample. All that is required is that the sub-populations be known and clearly defined. Each administrative subdivision or EA is assigned to the substratum which appears to contain most of its population.

In some cases, no information will be available at a central level on the number or size of the lowest level administrative units. In these cases it will be necessary to increase the number of stages of sampling, utilizing first only that information which is centrally available, and then moving into those subdivisions selected at the first stage to gather the additional information needed for sampling at subsequent stages. In this way researchers avoid having to gather information for non-selected regions. This process continues until the ultimate clusters containing the individual sample elements are identified. It is important to keep in mind, however, that the most important single influence on the precision of estimates obtained from a multistage sample is the number of cluster units included in the sample, not overall sample size. For this reason the sampling frame should be designed so that a minimum sample of 25-35 first-stage units within each stratum is feasible, and each substratum has at least two first-stage sample units.

To select the first-stage sample units to be interviewed proceed in the following manner:

1) Prepare a list of all first-stage sample units in the sub-stratum. In many cases it will prove useful to list these psus in serpentine fashion or in some other order that provides implicit stratification of the psus.

2) On the list note the sample population of each of the psus.

3) Construct a cumulative population listing, beginning with the sample population of the first psu on the list and adding to it the sample population of each subsequent psu. In this way the sample population of each psu occupies a unique and determinable interval within the cumulative list for the stratum.

4) When the listing is complete, divide the total number of sample elements in the stratum by the number of first-stage sample units desired in each stratum, preferably 25-30 per stratum. This yields the systematic sampling interval for selecting sample psus.

5) Then draw a random number between zero and the sampling interval. The first-stage sample unit containing that household number in its cumulative total becomes the first selected psu.

6) Proceed to select the remaining first-stage units by adding the systematic sampling interval to the first number and selecting each psu containing the designated sample element in its cumulative total. If one of the first-stage sample units is larger than the systematic sampling interval it will be selected more than once. In that case it should be allocated one second-stage sample unit for each time it is chosen. This process continues until the sample units at all stages except the last have been identified.

Once researchers have identified the ultimate clusters to be interviewed, enumerators must prepare an exhaustive list of sample units living in the selected clusters. If census enumeration areas are the first-stage sample units this requires that their boundaries first be identified on the ground. Enumerators will then pass from house to house or from local leader to local leader to obtain the names and addresses of households living within the

borders. Usually housing units rather than households are listed, and information is obtained to distinguish permanently occupied housing units (eligible units) from vacant, seasonal and other ineligible units. Adequate operational control is required to avoid duplication and omission and to ensure that the information obtained for each unit will be sufficient for the interviewer to locate the housing unit later if it is selected.

Frequently there will be a list of all household heads in an administrative unit at the headquarters of the unit. This is one advantage of using administrative units as a sampling frame. The boundaries are clear. Boundaries for census enumeration areas, on the other hand, sometimes cross political boundaries. Administrative lists, however, may not be complete or may be out of date.

The selection of the sample elements to be interviewed proceeds in the same fashion as at previous stages, except that the individual housing unit names are numbered and cumulated. The systematic sampling interval is determined by dividing the total number of housing units in the cluster by the size of sample to be chosen in the cluster, preferably 8-10 households per cluster. Those housing units with the serial numbers falling on the systematic sampling interval are then the ones to be interviewed for the survey.

In order to anticipate the likelihood of needing to replace certain elements in the sample it is desirable to draw more sample elements than one anticipates actually interviewing. This can most effectively be accomplished by increasing the initial desired sample size per cluster by 30-50% and then randomly drawing a subsample of the desired size. Alternatively, one can add or subtract one-half of the systematic sampling interval and take those households that fall on multiples of two times the interval as replacement households, selecting one at random as needed. Still, every effort should be made to avoid substitution and to retain the original sample as much as possible.

B. Weighting Procedures

In most real world situations pertaining to consumption surveys we have only a measure of the size of the population we want to study rather than the actual size. In order to maintain the representativeness of the sample we need to adjust the weights for the individual data observations for differences between the measure of size used to select a particular sample element, and the actual probability of selection as determined after we acquire more information about the sample units. At the same time, in order to obtain population totals, each element in the sample must be multiplied by the inverse of its selection probability, i.e. expanded, so as to reflect the fact that the sample represents a larger number of households in the population. These factors are all incorporated into a weight which is attached to each household record and applied to values that need to be extrapolated.

No matter how many stages are used in the sample selection process, the extrapolation factor or weight attached to a selected sample element is the reciprocal of that element's probability of selection. An element's probability of selection is simply the product of its probability of selection at each stage. Where sampling of all clusters within a substratum is with probability proportional to size (pps), and the number of sample units drawn at each stage after the first is constant across all elements in that stage, all elementary units within a substratum have an equal probability of selection and the sample is approximately self-weighting. Using notation similar to that used by Megill and Dauphin (1986) for the Haiti survey, this can be seen for a three-stage sample from the following:

$$P_{hijk} = \frac{n_h \quad M_{hi}}{M_h} \quad \frac{m_h \quad M_{hij}}{M_{hi}} \quad \frac{C_h}{M'_{hij}} \qquad (1)$$

where:

P_{hijk} = probability of selecting the k^{th} household in the j^{th} second-stage sample unit in the i^{th} first-stage sample unit in substratum h, constant for all households in j.

n_h = number of first-stage sample units selected from substratum h, variable from h to h.

M_{hi} = ex-ante estimate of the household population of the i^{th} first-stage sample unit in substratum h

M_h = ex-ante estimate of the household population of substratum h

m_h = number of second-stage units selected from the i^{th} first-stage sample unit of substratum h, constant for all i within substratum h.

M_{hij} = ex-ante estimate of the household population of the j^{th} second-stage sample unit in the i^{th} first-stage sample unit of substratum h

M'_{hij} = listed number of households in the j^{th} second-stage sample unit in the i^{th} first-stage unit in substratum h

C_h = number of households selected in the j^{th} second-stage sample unit in substratum h, constant for all j within substratum h.

The first term in equation (1) is the probability of selecting a given first-stage sample unit in substratum h. The second term is the probability of selecting a second stage unit given that the i^{th} first-stage unit has already been selected. The last term is the probability of selecting a household within the selected second-stage sample unit. Where the number of households actually listed in the selected second-stage sample unit, M'_{hij}, equals the ex-ante estimate of the household population of the unit, M_{hij}, equation (1) reduces to:

$$P_{hijk} = \frac{n_h \, m_h \, C_h}{M_h} \qquad (2)$$

This means that if a constant number of second-stage units are selected within each selected first-stage unit, and a constant number of third-stage units are selected within each selected second-stage unit, using pps in the first two stages and equal probability at the last stage, each household within the substratum has an equal probability of selection. The sample is, consequently, self-weighting. When the factors M'_{hij} and M_{hij} are not equal but are close, the sample is only approximately self-weighting. Where a two-stage sample rather than a three-stage sample is used, the term m_h drops out and each unit still has an equal probability of selection. In both cases if first-stage sample elements are allocated between substrata with

pps, and the m_h and c_h are equal between those substrata, the sample is self-weighting within a stratum. The same is true for the main strata. However, it is not usually advantageous to make a proportional allocation at that level because of variability in the data and substantial differences in population between the various strata, especially between urban and rural and high and low income strata.

Where there is reason to believe that there is greater variability between strata, obtaining the most cost-effective estimates of population estimates requires optimal allocation of sample units between the strata. Essentially, this means allocating sample observations to the various strata as a function of variability within each stratum and the cost of interviewing a sample unit in that stratum. At the same time, when separate estimates with a minimum precision are needed for each stratum, a minimum sample size would have to determined for each stratum on its own. Merely allocating the overall sample to the strata, whether by optimum allocation or proportional allocation, would not guarantee the required level of precision.

In practice M'_{hij} will not equal M_{hij}. Population will have changed since the last measure no matter how recent it is. Moreover, when a listed housing unit is visited, the enumerator will sometimes find more or fewer households than was anticipated on the basis of the household listing. To be rigorous, the weights attached to those households actually interviewed within those last stage sample units need to be adjusted to reflect the modified probability of selection.

In the case of finding additional households where there was thought to be only one, one of the households should be selected for the interview and its weight multiplied by the number of household units identified. This corrects for the fact that there are more households of this type in the population than was foreseen in the original sample.

Where a listed household no longer exists (is vacant, destroyed or is seasonal) at the time of the interview and a new household has been substituted in order to maintain the level of precision desired for the estimates, the weights of <u>all</u> of the other households interviewed within the second-stage unit need to be adjusted. The adjustment would reflect the fact that there are fewer households in that cluster than was anticipated. The adjustment factor would be:

$$\frac{C_h - d_{hij}}{C_h} \qquad (3)$$

where:

d_{hij} = number of ineligible sample households in the j^{th} second-stage sample unit in the i^{th} first-stage sample unit in substratum h

The total weight of each household is then its initial probability of selection multiplied by the appropriate adjustment factors. In cases where the ineligible households are not replaced in the sample, such a weight reduction is not necessary.

Other adjustments in the weights may be necessary to allow for non-response and consolidation of households. The non-response adjustment is required when fewer than C_h households are interviewed in the cluster despite substitution efforts. The adjustment factor is simply C_h divided by the number of completed interviews within the cluster j and is multiplied by the weight of each interviewed household in the cluster. Since this adjustment can introduce bias it should not be used when there are more than 25-30% missing units.

The adjustment for consolidation is necessary when an interviewed household's probability of being interviewed was augmented because it occupies more than one housing unit. This occurs when one household occupies the same space that was occupied by two different households at the time of listing. As a consequence,

its probability of selection has doubled since it will be picked if either one of the two originally listed households are selected. It is therefore necessary to multiply the weight of this interviewed household by one-half.

Unless there is some reason to suspect that significant demographic changes have occurred since the data used to construct the sampling frame were originally collected, the bias introduced by ignoring the adjustments for missing and unanticipated sample units will sometimes not be great in relation to total sampling and non-sampling errors. In such cases it would be reasonable to assume that such phenomena occur at random throughout the population. Simply replace a missing household with one of the substitutes. Where there is more than the one sample unit, choose one for the interview and proceed as though nothing occurred In both cases the sample size will remain the same constant size, preserving the self-weighting sample and keeping work schedules intact. At the same time, make note of these occurrences in the event that a subsequent tabulation by computer permits use of variable weights. Variable weights can be attached at that time. Of course, users of the data should be informed of any such shortcuts that could compromise the usefulness of the data for their particular analytical purposes. Where a data base is expected to be used by a number of analysts for as yet unidentified purposes, it would be best to make every effort to calculate the correct weights.

In addition to weights associated with sampling probabilities for the selected elementary sample units, each elementary unit requires a weight to extrapolate recorded expenditures to the appropriate time interval. This weight will depend on the data collection methodology employed for the survey. To estimate annual consumption and expenditures, the extrapolation factor would be 365/r where r is the reference period covered by the survey (in days). This factor will normally vary for different types of expenditures within the same survey, as well as for different sur-

veys. For single interview approaches r would normally be rather small.

C. Statistical Inference

When done *via* a cumulative population listing, systematic sampling of psus is equivalent to sampling without replacement. Generally speaking, sampling without replacement is more precise than sampling with replacement. When the sampling fraction is small, the chance of the same unit being selected twice is also small. In this case sampling with replacement is practically equivalent to sampling without replacement. Frequently, however, especially when using sampling frames based on administrative units, the first-stage sampling fraction will be quite high. In this case it will be necessary to introduce the finite population correction factor (fpc) to avoid overestimating the variance when sampling is done without replacement.[1]

To maintain the widest range of flexibility for estimating parameters and their variance we will assume that the sample is only approximately self-weighting. This means that estimates will be biased in proportion to the error in the measures of size. However the variance of the estimates proposed here will often be substantially smaller than with many unbiased estimates under the same circumstances. Moreover, the bias in the estimates will not be large unless the measures of size are substantially in error. This will not usually be the case with census or administrative data.

Turning first to estimating totals, means and proportions, again using notation similar to Megill and Dauphin (1986), the survey estimate of a total would be:

---

[1] The finite population correction factor for first-stage units is $1 - n/N$, where n is the number of first-stage units sampled and N is the total number of first-stage units in the population. The fpc factor is always less than one and always reduces the variance.

$$X_A = \sum_{k=A} W_k X_k \qquad (4)$$

where:

X_A = total of variable X for group A

A = subset of records belonging to group A

W_k = final weight for the k^{th} record

X_k = value of variable X for the k^{th} record

To calculate means and proportions we use the ratio estimate Y_A/X_A, where Y_A and X_A are the corresponding total estimates. According to Megill and Dâuphin (1986):

> "...Means and proportions are special types of ratios. In the case of a mean, the variable X in this ratio would be equal to 1 for each record in group A, so that the denominator would equal the sum of the weights for group A. In the case of a proportion, the variable Y in the numerator of the ratio would be either 1 or 0 depending on the presence or absence of a specified characteristic." (p.7)

The term W_k in equation (4) can be a true final weight, i.e. one that includes adjustment for missing or additional households, or a simple extrapolation factor based on the initial probability of selection for household k. In the former case, X_A is an asymptotically unbiased estimate of the population total of X. In the latter case X_A is biased and the bias is proportional to to the errors in the W_k. As mentioned previously, unless there is reason to believe that there have been substantial population shifts since the population figures used in the sampling frame

were gathered, this bias is probably not quantitatively substantial in relation to non-sampling error and can be ignored. This permits the data to be tabulated as a simple random sample with extrapolation factors applied only once, to the sample totals.

Estimating variances is considerably more difficult. The Haiti survey used SUPER CARP, a mainframe generalized variance software package developed at Iowa State University, and the corresponding microcomputer version, PC CARP. SUPER CARP provides for the calculation of variances for estimates of totals, means, proportions and other ratios. Apparently, the software does not provide a finite population correction factor for multistage sampling with first-stage selection probabilities proportional to size. Whether or not this is a problem will depend on the proportion of first-stage units included in the sample. In most circumstances this will be less than 10% so that ignoring the fpc factor will not make a great difference in the results.

CLUSTER (Verma and Pierce,1978) is another program that estimates variances of multistage samples and cluster samples. Annex A provides equations for calculating variances from two-stage and three-stage samples selected with probabilities proportional to a measure of size for those who do not have access to a canned program for calculating variances.

VII. CONCLUSION AND RECOMMENDATIONS

In most countries data on census enumeration areas are readily available. Census enumeration areas are uniform in size, have a relatively known population, are small enough that they can be easily grouped into relatively homogeneous strata and can usually be covered by one enumerator on foot. Where enumeration areas are selected with probability proportional to size and the sample is self-weighting, relatively simple, two-stage estimating procedures apply. In this case variances can be calculated with a hand held calculator.

The principal drawback of using EAs as sample units is the time involved in obtaining maps or definitions of their boundaries and obtaining the list of elementary sample units in the selected EAs. Since EA boundaries do not always follow neighborhood or political boundaries, special care is required to insure that the unit is properly defined. Some countries maintain detailed maps and/or clear definitions of the boundaries. Many, however, do not. Even where they do, they do not maintain an up-to-date list of households. Consequently, it is virtually always necessary to do an on-the-ground listing of elementary sample units before being able to proceed with sampling within the selected EAs. Before deciding to use census enumeration areas as first-stage sample units in a rapid appraisal approach, researchers should verify that sufficient information is available to ensure quick and proper identification on the ground.

In contrast to census enumeration areas, good population data are almost always available on administrative units down to the lowest levels. Lists of households within the lower level units are usually available at local government headquarters, often eliminating the need for an on-ground listing.[1] But administrative units have their downside too. The size and number of the lower level units likely to be used as ultimate cluster sampling units is not usually known in the capital city or in regional capitals. This will normally necessitate the use of three or more stages of sampling and require more complicated estimating procedures.

Ideally, the first-stage sample units in an administrative type sampling frame would be the lowest level administrative unit for which population data are available at central levels. The lowest unit for which the total number of units in the overall population is known would then serve as the ultimate cluster.

[1] Administrative lists frequently are incomplete. But on-ground listing is not always better. Outside enumerators overlook sample units too, or rely on the recall of others who do the same.

This will insure that data needed for the finite population correction factor will be available. Sampling below that level would be used only to provide unbiased estimates of the ultimate cluster values used to make population estimates. Individual household data would still be available for analyses requiring such data.

This report presents procedures and estimators for using both census enumeration areas and administrative units as a sampling frame for gathering household consumption and expenditure data. Other approaches could be used as well. Each approach will have a separate set of estimators. Which one to use is a question of the cost-effectiveness of the estimates. Stratified, multistage sampling with probabilities proportional to size in the context of a self-weighting sample normally provides very cost-effective estimates. Sampling with probabilities proportional to a measure of estimated size rather than actual size, the normal situation under field conditions, provides estimates that are almost as efficient, but somewhat biased. As indicated previously, the amount of bias will not be large unless the measure of size is substantially in error. Even this bias can be reduced if individual sample element weights are adjusted to reflect the divergence between the measure of size actually used and the actual size found at the time of listing. Before such an adjustment makes sense, however, researchers must satisfy themselves that such differences are occurring because of population dynamics rather than because of an imprecise definition of sampling clusters at two different points in time. In practice, this will not be easy.

As this report has shown, it takes little additional time to develop and utilize a statistically rigorous sampling methodology. The major cost is in terms of additional visits to a larger number of primary sample units and the political and logistical preparation that this requires. This is usually an insignificant cost in relation to total survey costs. Yet a serious effort to develop and apply an efficient and rigorous sampling methodology greatly

expands the analytical usefulness and the potential policy relevance of consumption and expenditure surveys.

The specific recommendations of this report include the following:

1) The Nutrition Economics Group should finance a combination consumption and expenditure survey that includes a rigorous examination of the design effect and the effect on data quality of various interview frequencies aimed at developing estimates valid for population groups. The study should include a comparison of single interview, quarterly interview and monthly interview approaches in an urban and a rural environment.

2) Consumption/expenditure surveys should use statistically valid, self-weighting, stratified, multistage probability sampling wherever possible. This will greatly simplify calculation of variances, reduce costs and help contain non-sampling errors. The sampling methodology should include the following:

- Use of census enumeration areas or the lowest level administrative units for which population data are centrally available as the primary sample units.
- Grouping of primary sample units into strata of relatively homogeneous sample units.
- Structuring the sample so that there will be a minimum of 25-30 primary sample units selected within each stratum.
- Selecting sample clusters at all stages with probability proportional to estimated size. At all stages except the first draw a constant number of sample units within each stage.
- The size of the sample drawn at the last stage of sampling should be 8-10 units per cluster. A larger sample size per ultimate cluster would be called for where the underlying variability in the population with respect to the parameter to be measured is large. Gen-

erally, however, a larger number of primary sample units is preferable to a larger number of sample elements per cluster for a given sample size.

BIBLIOGRAPHY

Ahlers, Theodore H., et. al. 1983. "Le Secteur Rizicole a Madagascar: Etude du Secteur Rizicole, Rapport de la Premiere Phase". Service de Coordination du Secteur Riz, Ministere de la Production Agricole at de la Reforme Agraire and Associates for International Development. Somerville.

Cochran, William G. 1953. Sampling Techniques. John Wiley and Sons, Inc. New York.

Freudeheim, Nancy Johnson and Robert Wardrop; 1987; "Misclassification of Nutrient Intake of Individuals and Groups Using One, Two, Three and Seven Day Food Records"; American Journal of Epidimiology, Vol. 126, No. 4.

Hansen, Morris H., William N Hurwitz and William G. Madow. 1953. Sample Survey Methods And Theory. John Wiley and Sons, Inc. New York.

Hiemstra, Stephen J. 1987. "Urban Food Consumption Patterns And National Food Policy In Liberia: Methodology And Evaluation". Nutrition Economics Group, USDA and Purdue University.

Megill, David J. and Marjorie M. Dauphin. 1986. "Estimation Procedures For The Haiti Household Expenditure And Consumption Surveys: Weighting Procedures and Calculation of Variances". U.S. Bureau of the Census. Washington,D.C.

Megill, David J. and Marjorie M. Dauphin. 1985. "Preliminary Recommendations for Sample Design of Household Expenditure Survey in Haiti". U.S. Bureau of the Census. Washington, D.C.

Senauer, Ben. 1987. "Rapid Appraisal: The Survey Instrument". Report Prepared for the Nutrition Economics Group, USDA.

Spurr, William A. and Charles P. Bonnini. 1967. Statistical Analysis For Business Decisions. Richard D. Irwin. Homewood.

Verma, Vijay and Mick Pearce; 1978; "User's Manual For CLUSTERS: A Package Program For Computation Of Sampling Errors For Clustered Samples"; International Statistical Institute, World Fertility Survey; London.

Yates, Frank. 1960. Sampling Methods For Censuses And Surveys. Hafner Publishing Company. New York.

Zalla, Thomas M. 1982. "Economic and Technical Aspects of Smallholder Milk Production in Northern Tanzania." Unpublished Ph.D. dissertation. Department of Agricultural Economics, Michigan State University. East Lansing.

Zalla, Tom, 1987. "Toward Rapid Appraisal of Consumption and Expenditure Data". Report prepared for the Nutrition Economics Group, USDA.

ANNEX A

VARIANCE FORMULAS FOR A STRATIFIED, MULTI-STAGE SAMPLE WITH SELECTION PROBABILITIES PROPORTIONAL TO A MEASURE OF SIZE

Using an ultimate cluster type of estimator, and including the FPC factor, equation (5) gives an estimate, slightly biased, of the variance of a total (equation 4, page 29) for a self-weighting (within strata), three stage sample selected with probabilities proportional to a measure of size at the first two stages:

$$\mathrm{Var}\ (\hat{X}) = \sum_{h=1}^{l} W_h^2 \left(1 - \frac{n}{Q_h}\right) \frac{n}{n-1} \sum_{i=1}^{n_h} \sum_{j=1}^{m_h} \left(\hat{X}_{hij} - \frac{\hat{X}_h}{n}\right)^2, \qquad (5)$$

where:

h, i, j, and k refer to substrata, first-stage, second-stage and third-stage sample values;

M_h, n_h and m_h are defined as for equation (1)

l = number of substrata in the population

W_h = M_h/M = the weight of substrata h in the total population where M is the household population of all substrata combined

Q_h = number of second-stage units (ultimate clusters) in substratum h

n = $n_h\ m_h$ = number of second-stage units (ultimate clusters) selected in substratum h

$1 - \frac{n}{Q_h}$ = the finite population correction factor for second-stage sample units (ultimate clusters)

$$\hat{X}_{hij} = \sum_{k=1}^{c_h} W_{hijk} X_{hijk}$$ = weighted total of variable X for the j^{th} second-stage unit in substratum h

where:

W_{hijk} = final weight for the k^{th} sample unit in the j^{th} second-stage unit of the i^{th} first-stage unit in substratum h

X_{hijk} = value of variable X for the k^{th} sample unit in the j^{th} second-stage unit in the i^{th} first-stage unit in substratum h

$$\hat{X}_h = \sum_{i=1}^{n_h} \sum_{j=1}^{m_h} \hat{X}_{hij}$$ = weighted total of X for substratum h

$$\hat{X} = \sum_{h=1}^{l} \hat{X}_h$$ = weighted total of X for the population

The reader will notice that (5) is just the weighted sum of the variances of the n second-stage weighted sample totals in each of the l substrata. Where the number of sample units is constant within each cluster in a substratum, so that individual household weights[1] are also constant, it can be calculated manually without too much difficulty.

Where two stage sampling is all that is needed the estimated variance of a total would be:

$$\mathrm{Var}\ (\hat{X}) = \sum_{h=1}^{l} W_h^2 \left(1 - \frac{n_h}{N}\right) \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left(\hat{X}_{hi} - \frac{\hat{X}_h}{n_h}\right)^2, \qquad (6)$$

[1]Assuming adjustments for missing or duplicate households are ignored.

where:

N = number of first-stage units (ultimate clusters) in substratum h

$(1 - \frac{n_h}{N})$ = the finite population correction factor for first-stage sample units

$\hat{X}_{hi} = \sum_{j=1}^{C_h} W_{hij} X_{hij}$ = weighted total of variable X for the i^{th} first-stage unit in substratum h

$\hat{X}_h = \sum_{i=1}^{n_h} \hat{X}_{hi}$ = weighted total of X for substratum h

For the variance of means and ratios SUPER CARP provides estimates that are rather complex. Equation (7) provides a sometimes more biased but easier to calculate estimate for a three stage sample selected with pps:

$$Var\ (\bar{X}) = \sum_{h=1}^{1} W_h^2 \ (1 - \frac{n}{Q_h}) \ \frac{n}{n-1} \ \sum_{i=1}^{n_h} \ \sum_{j=1}^{m_h} (\bar{X}_{hij} - \bar{X}_h)^2, \qquad (7)$$

where:

$\bar{X}_{hij} = \dfrac{\sum_{k=1}^{C_h} W_{hijk} X_{hijk}}{\sum_{k=1}^{C_h} W_{hijk}}$ = weighted mean of the j^{th} second-stage unit in the i^{th} first-stage unit of substratum h

$\bar{X}_h = \dfrac{\sum_{i=1}^{n_h} \sum_{j=1}^{m_h} \bar{X}_{hij}}{n_h \ m_h}$ = Weighted mean for substratum h

Equation (8) gives an estimate for the variance of a mean for the same kind of two-stage sample:

$$\mathrm{Var}\ (\bar{X}) = \sum_{h=1}^{l} W_h^2 \left(1 - \frac{n_h}{N}\right) \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (\bar{X}_{hi} - \bar{X}_h)^2, \qquad (8)$$

where:

$$\bar{X}_{hi} = \frac{\sum_{j=1}^{C_h} W_{hij}\ X_{hij}}{\sum_{j=1}^{C_h} W_{hij}}$$

= weighted mean of variable X for the i^{th} first-stage sample unit in substratum h

$$\bar{X}_h = \frac{\sum_{i=1}^{n_h} \bar{X}_{hi}}{n_h}$$

= weighted mean of variable X for substratum h

ANNEX B: GLOSSARY

Accuracy: In a statistical sense accuracy refers to how closely an estimated value approximates the true (and unknown) population value. Accuracy includes the effect of both precision and bias.

Area Sampling: A method of sampling in which sample elements are defined as geographical units. Normally, the areas to be sampled are defined so as to be approximately equal in size.

Asymptotically Unbiased: A characteristic of an estimator such that it becomes less biased as the sample size approaches the population size.

Bias: A persistent tendency to err in a particular direction.

Cluster: A group of sample units that are closely associated in space.

Cumulative Population Listing: A method of listing sample units such that the population of each successive sample unit is added to the total of previous sample units, and the interval occupied by each sample unit is noted. Thus, each sample unit occupies a unique interval on the list of total sample units.

Design Effect: A measure of the extent to which clustering sample observations increases the size of a sample relative to a simple random sample that is required to obtain the same level of precision as a simple random sample.

Finite Population Correction Factor (FPC): A number that reduces the estimate of a variance according to the proportion of population units included in the sample. The greater the proportion of the population included in the sample, the smaller is the FPC and the resulting estimate of the variance. The FPC always reduces the estimated variance, though not by much when the population is large relative to the sample.

First-Stage Sample Unit: The population unit selected at the first stage of sampling. First-stage sample units are generally administrative units, census enumeration areas or geographic units for which the population is known but which is too large to cover in its entirety.

Intraclass Correlation: The similarity between elements in the population that are clustered together. Usually, elements in close proximity to each other are more similar than elements which are more dispersed. This reduces the precision of a sample of a given size.

Non-Sampling Errors: Errors that result from sources other than sampling. Non-sampling error includes error introduced by such things as poorly supervised field staff, poorly worded questions, errors in data recording, editing and analysis, respondent fatigue or deception and similar matters.

Precision: The extent to which a sample reproduces the results that would be obtained if we took a complete census using the same methods of measurement, etc.

Primary Sample Unit: See first-stage sample unit.

Proportional Allocation: The process of distributing a sample across subgroups of a population in proportion to the number of sample elements in the subgroup.

Random Number: A number selected in such a way that all numbers in a specified range have an equal probability of being selected. Normally we use a computer generated random number table to select random numbers when the sample population is large.

Sample Elements: The ultimate units from which we collect data.

Sample Frame: A list of sample elements or sample units containing the sample elements.

Sample Units: The entities which we select in a sample at any stage. Only the last stage sample units will be sample elements.

Sampling Error: The random or chance error that occurs when we take a sample rather than a census of the whole population. Sampling error can be reduced by increasing sample size.

Self-Weighting Sample: A sample that is designed in such a way that the unweighted mean of the sample gives an unbiased estimate of the population mean.

Serpentine Listing: A process of listing sample units that criss-crosses an area in such a way as to ensure a good dispersion of sample units across space when utilizing a systematic sampling interval.

Simple Random Sample: A sample in which all sample elements have an equal probability of being selected and are drawn directly from the sample population without clustering or utilizing more than one stage of sampling.

Standard Deviation: A measure of the dispersion of population (sample) elements about their mean. The standard deviation is the square root of the variance.

Standard Error: The standard deviation of the sample estimate divided by the square root of the sample size. The standard error measures the extent to which the sample estimate differs from the true population estimate. It approaches zero as the size of the sample approaches the population size.

Stratified Sample: A sample from a population that has first been divided into groups of known size. Stratifying populations according to their similarity with respect to the things being measured usually increases the precision of sample estimates for a given size of sample.

Stratum: A group of population elements that share similar characteristics with respect to the objectives of a sample survey.

Systematic Sampling: A method of sampling in which a fixed sampling interval is applied to a list of population units in order to identify the specific units to be included in the sample. The fixed sampling interval is called the systematic sampling interval.

Two-Stage (Multistage) Sample: A sample drawn in more than one stage. Normally all but the last-stage sample units are made up of groups of sample elements that are too numerous to survey but for which the population is known. This reduces the number of households that ultimately have to be listed in order to draw the sample elements to be interviewed. Multistage sampling usually reduces survey costs.

Ultimate Cluster: The lowest cluster or sample unit from which sample elements are selected.

Variance: A measure of the dispersion of population (sample) elements about their mean. The variance is a measure of the squared difference of the population (sample) elements from the population (sample) mean.

Weight (extrapolation weight, extrapolation factor, expansion factor): A number attached to a sample observation equal to the inverse of that sample element's probability of selection. This number, relative to the total of all weights, reflects the share of the population total represented by the sample element to which it is attached.